

Two-way ANOVA for scalar trajectories, with experimental evidence of non-phasic interactions

Todd C. Pataky¹, Jos Vanrenterghem², and Mark A. Robinson²

¹Department of Bioengineering, Shinshu University, Japan

²Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, UK

December 2, 2014

Abstract

Kinematic and force trajectories are often normalized in time, with mean and variance summary statistic trajectories reported. It has been shown elsewhere, for simple one-factor experiments, that statistical testing can be conducted directly on those summary statistic trajectories using Random Field Theory (RFT). This technical note describes how RFT extends to two-factor designs, and how bizarre “non-phasic interactions” can occur in multi-factor experiments. We reanalyzed a public dataset detailing stance phase knee flexion during walking in (a) patellofemoral pain vs. controls, and (b) females vs. males using both a full model (with interaction effect) and a main-effects-only model. In both models the main effect of PAIN failed to reach significance at $\alpha=0.05$. The main effect of GENDER reached significance over 5–40% stance ($p=0.0005$), but only for the full model. The interaction effect (in the full model) reached significance over 0–15% of stance ($p=0.030$), and resulted from greater flexion in females but decreased flexion in males in PFP vs. controls. Thus there was a non-phasic interaction, in which a non-significant interaction (over 20–40% stance) suppressed the main effect of GENDER. Similarly, if we had only analyzed 20–40% stance, we would have committed Type II error by failing to reject the null PAIN-GENDER interaction hypothesis. The possible presence of non-phasic interactions implies that trajectory analyses must be conducted at the whole-trajectory level, because a failure to do so will generally miss non-phasic interactions if present.

Keywords: *kinematics; statistical parametric mapping; random field theory; time series analysis*

1 Introduction

Biomechanical processes are often summarized using one-dimensional trajectories which usually represent kinematics or forces, and which have been registered (Sadeghi et al., 2003) to some homologous temporal domain, often by linearly interpolating between 0% and 100% time. This paper pertains to analysis of such data.

If an investigator has no specific *a priori* hypothesis where in the range 0–100% a kinematic or force effect is expected to emerge, then by definition the null hypothesis implicitly pertains to the entire trajectory (Pataky et al., 2013). This null hypothesis of “trajectory equivalence” is valid, but to test it objectively one must consider the behavior of random data under that null hypothesis. In particular, from a classical hypothesis testing perspective, one must compute an α -defined critical threshold above which random data would traverse in only $\alpha\%$ of many repeated experiments.

Random field theory (RFT) (Adler and Taylor, 2007) describes the behavior of smooth n -dimensional Gaussian continua, and in particular the probability that they will produce test statistic continua which exceed arbitrary thresholds in arbitrary experiments. RFT has been used in applied form most widely in the neuroimaging literature (Friston et al., 2007) and has also been applied to smooth kinematic/force trajectories in simple one-factor experimental designs (Pataky et al., 2013), but the biomechanical implications of trajectory-level two-way ANOVA have not yet been explored. The purposes of this Technical Note were: (1) to demonstrate trajectory-level two-way ANOVA, and (2) to explore the biomechanical implications of including/excluding interaction terms in statistical models.

2 Methods

2.1 Data

A public dataset detailing stance-phase knee flexion during walking in subjects with patellofemoral pain (PFP) (Besier et al., 2009) was reanalyzed (Fig.1, Table 1). The dataset consisted of 41 subjects, including: 8 control females, 7 control males, 16 PFP females and 10 PFP males. We subsequently refer to this 2×2 design using the factor labels PAIN (PFP vs. controls) and GENDER (females vs. males). The public dataset was linearly interpolated to 100 time points over stance phase and contained one mean trajectory per subject, as estimated from at least three trials of self-paced walking.

2.2 Statistical analysis

To assess the biomechanical implications of interaction effects, we analyzed the data using both a full two-way ANOVA (with interaction):

$$y_{ijkq} = (\tau_{\text{pain}})_{iq} + (\tau_{\text{gender}})_{jq} + (\tau_{\text{pain,gender}})_{ijq} + \varepsilon_{ijkq} \quad (1)$$

and a main-effects only model (without interaction):

$$y_{ijkq} = (\tau_{\text{pain}})_{iq} + (\tau_{\text{gender}})_{jq} + \varepsilon_{ijkq} \quad (2)$$

where y_{ijkq} is the experimental observation for the k th subject of the i th level of PAIN, j th level of GENDER and q th point in time, $(\tau_{\text{pain}})_{iq}$ and $(\tau_{\text{B}})_{jq}$ are group means at the q th point in time for the i th level of PAIN and the j th level of GENDER, respectively, and $(\tau_{\text{pain,gender}})_{ijq}$ is the interaction term modeling possibly different effects of GENDER on the i th level of PAIN. The ε_{ijkq} term represents model residuals. Note that Eqns.1&2 model three and two factors, respectively, and that there are therefore three and two F statistics, respectively, to compute.

We followed typical two-way ANOVA procedures to calculate F values separately at each time point q , thereby forming F statistic trajectories (see Supplementary Material). We also corrected for potential non-sphericity (i.e. potentially unequal variance across PAIN/GENDER levels) using restricted maximum likelihood estimates of the degrees of freedom (Friston et al., 2007).

We next conducted classical hypothesis testing at a Type I error rate at $\alpha=0.05$. Noting that RFT assumes that the residuals ε_{ijkq} are smooth, Gaussian random fields, and that this assumption has been validated elsewhere for biomechanical trajectories (Pataky et al., 2014), we used RFT’s analytical descriptions of smooth Gaussian field behavior to compute the critical threshold F^* that identically smooth Gaussian fields would reach in only $\alpha\%$ of identical, repeated experiments. On this basis an F trajectory which exceeds F^* leads to null hypothesis rejection.

Last, we computed precise probability values for supra-threshold clusters in a similar manner. Briefly, a thresholded F trajectory generally contains a collection of supra-threshold trajectory segments (or ‘clusters’), and RFT yields analytical solutions for the probability that a cluster of a particular extent (i.e. temporal length) would be produced at the particular threshold (Friston et al., 2007). By definition, a cluster which just touches the threshold F^* has a probability value of α , and p values decrease as cluster extents increase.

All analyses were implemented in Python 2.7 using Canopy 1.3 (Enthought Inc., Austin, USA). All

computational details are available in our open-source software at www.spm1d.org.

3 Results

Although joint angle trajectories were quite variable across subjects (Fig.1), mean trajectories (Fig.2) exhibited notable qualitative differences. In particular, in controls the mean male knee angle was greater than the mean female knee angle over the entire stance phase (Fig.2c), but in the PFP group the mean male and female trajectories were quite similar (Fig.2d). This implies that PFP tended to produce different effects in males vs. females, and this qualitative inference can be observed in Fig.2a,b.

Results for the full statistical model (Eqn.1, Fig.3) found that, while the main effect of PAIN failed to reach significance, the main effect of GENDER reached significance over 5–40% stance ($p=0.0005$) and the interaction effect also reached significance over 0–15% stance ($p=0.030$). In contrast, neither the main effect of PAIN nor the main effect of GENDER reached significance for the main-effects-only model (Eqn.2, Fig.4). Note, in particular, a “non-phasic” interaction effect: the interaction effect, which spans only 0–15% stance in Fig.3c, does not temporally overlap with the second peak in the main effect of GENDER (Fig.3b, 20–40% stance) yet this 20–40% main effect of GENDER was absent in the second model’s results (Fig.4b).

4 Discussion

The key new results were: (1) a demonstration of trajectory-level two-way ANOVA, and (2) an identification of non-phasic interaction possibilities in biomechanical trajectories. The former is important in the context of the Biomechanics literature, partially because only simple one-factor experiments have been demonstrated previously (Pataky et al., 2014), but more importantly because it shows that we can conduct a single statistical test — which simultaneously tests all trajectories at all points in time — for arbitrarily complex experiments. By conducting only a single test, we maximize statistical power because we don’t have to correct for multiple tests conducted on multiple scalars extracted from the single trajectories.

The latter result — non-phasic interaction — has important implications for all experiments involving scalar/vector trajectories. First, and most simply, if we had not modeled the interaction, we would have failed to reject the null hypothesis regarding the main effect of GENDER (as in Fig.4). This reiterates basic two-way ANOVA theory, and is applicable to all analyses, whether trajectory-level or not: a lack of a main effect in a particular factor (i.e. GENDER, Fig.4b) does not justify pooling across levels of that factor because interaction effects may hide inter-level differences (Fig.3c).

Much less trivially, the results also show that a significant interaction in one trajectory phase (Fig.3c) can appear to amplify a main effect in a separate phase (Fig.3b). We call this a “non-phasic interaction” and its implications are important: had we decided — in an *ad hoc* manner — to only analyze data in the vicinity of the first knee flexion peak (25–30% stance) (Fig.1) we would have found that there was a main effect of GENDER (Fig.3b) but not an interaction effect (Fig.3c). We would thus unjustifiably conclude that PFP does not affect males and females differently. In other words, interaction effects, which are themselves primary results, both vary in time and can alter the main effects in a time-dependent manner. non-phasic interactions therefore provide strong support for the notion that objective testing of hypotheses pertaining to whole trajectories requires trajectory-level techniques.

Although the biomechanical meaning of the observed non-phasic interaction is unclear, it is biomechanically clear that early-phase behaviors can produce cumulative effects on later phases (Richter et al., 2014). It is therefore possible that the presently observed non-phasic interaction relates to late-stance trajectory convergence (Fig.1). Regardless, the precise interpretation is scientifically irrelevant; Fig.3c shows that the no-interaction null hypothesis is correctly rejected for this dataset, and to scientifically probe its biomechanical meaning an investigator must derive a relevant hypothesis to test in a future experiment.

The main limitation of the present RFT approach is that it assumes homologous data registration. This is potentially problematic because apparently homologous events like local maxima may not be aligned precisely in time (Fig.1), and therefore non-linear registration, by definition, reduces trajectory variance (Sadeghi et al., 2003). Future studies should consider sensitivity of RFT results to registration particulars and to potential mis-registrations. Nevertheless, registration’s limitations are not unique to RFT-based inference; all analysis techniques require homologous data comparison, and mis-registration could affect all trajectory analyses. In particular, the common approach of extracting scalars from particular trajectory regions does not guarantee homologous data comparison.

In summary, this study has shown that classical hypothesis testing can be conducted at the whole-trajectory level for two-way ANOVA designs, and by implication, for arbitrarily complex experimental designs, using an α -based RFT critical threshold. More importantly, this study has also demonstrated that non-phasic interactions can exist in scalar trajectory datasets. Further investigations on independent datasets are required to determine the likelihood of observing such effects in general datasets. While 0D analysis of 1D data generally yields invalid statistical conclusions, this does not imply that clinical/biomechanical interpretations of 0D results are also invalid. For maximum objectivity, 1D analysis should be conducted when one’s *a priori* hypothesis does not pertain to a specific temporal instant or region.

Conflict of Interest

The authors report no conflict of interest, financial or otherwise.

References

- Adler, R. J. and Taylor, J. E. 2007. *Random Fields and Geometry*, Springer-Verlag, New York.
- Besier, T. F., Fredericson, M., Gold, G. E., Beaupre, G. S., and Delp, S. L. 2009. Knee muscle forces during walking and running in patellofemoral pain patients and pain-free controls, *Journal of Biomechanics* **42(7)**, 898–905, data: <https://simtk.org/home/muscleforces>.
- Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and Penny, W. D. 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Elsevier/Academic Press, Amsterdam.
- Pataky, T. C., Robinson, M. A., and Vanrenterghem, J. 2013. Vector field statistical analysis of kinematic and force trajectories., *Journal of Biomechanics* **46(14)**, 2394–2401.
- Pataky, T. C., Vanrenterghem, J., and Robinson, M. A. 2014. Parametric and non-parametric confidence intervals and hypothesis tests regarding scalar trajectories., *Journal of Biomechanics* *in review*.
- Richter, C., O’Connor, N. E., Marshall, B., and Moran, K. 2014. Analysis of characterizing phases on waveforms – an application to vertical jumps, *Journal of Applied Biomechanics*, in press.
- Sadeghi, H., Mathieu, P. A., Sadeghi, S., and Labelle, H. 2003. Continuous curve registration as an intertrial gait variability reduction technique, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **11(1)**, 24–30.

Table 1: Subject details (means \pm SD); replicated from Besier et al. (2009).

	Controls		Patellofemoral pain	
	Males ($n=8$)	Females ($n=8$)	Males ($n=11$)	Females ($n=16$)
Age (years)	27.2 \pm 3.0	28.8 \pm 4.7	30.5 \pm 4.5	28.7 \pm 4.6
Height (m)	1.79 \pm 0.07	1.66 \pm 0.05	1.78 \pm 0.09	1.68 \pm 0.06
Mass (kg)	74.2 \pm 4.2	58.3 \pm 4.6	72.4 \pm 12.5	62.7 \pm 10.0
Walking speed (m/s)	1.49 \pm 0.12	1.43 \pm 0.15	1.52 \pm 0.14	1.52 \pm 0.20

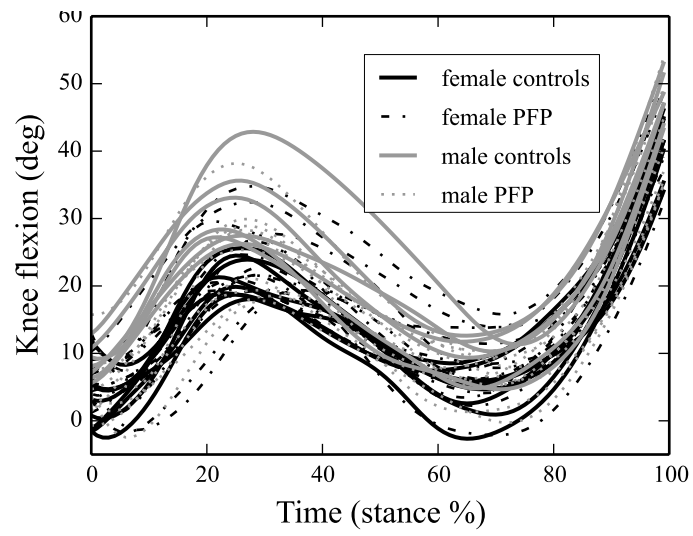


Figure 1. Sagittal plane knee kinematics, one trajectory per subject (from Besier et al., 2009).

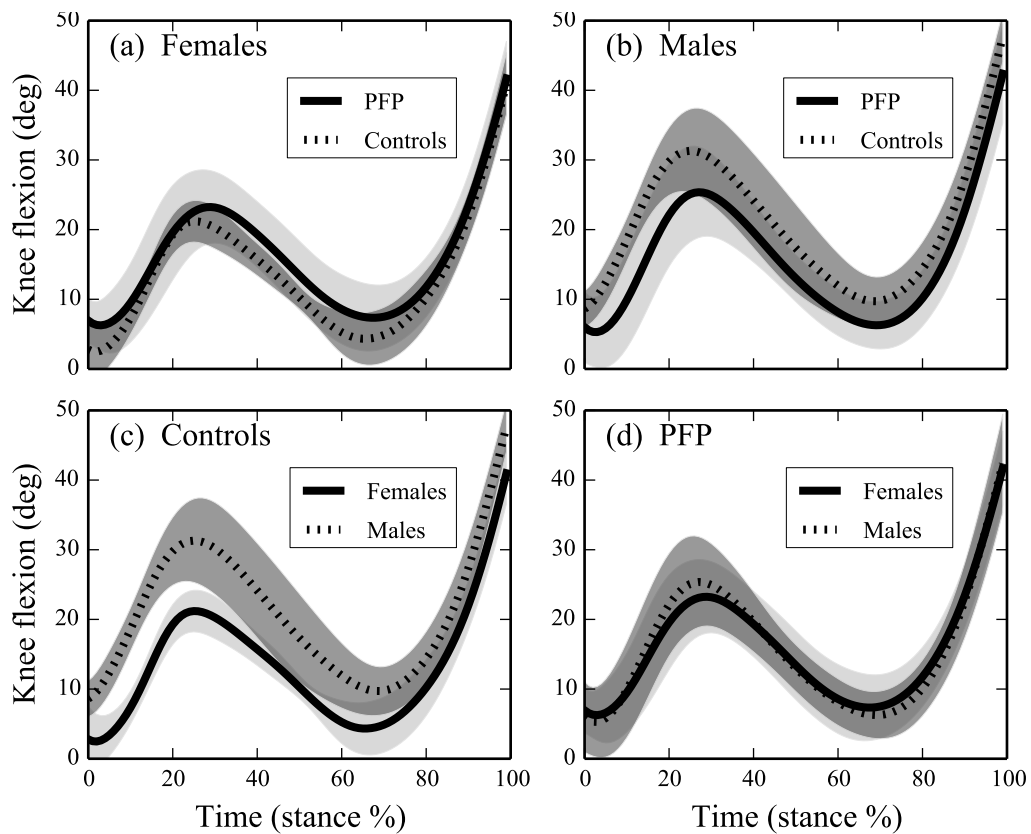


Figure 2. Group means with SD clouds. Top panels: PFP vs. Controls in (a) Females and (b) Males. Bottom panels: Females vs. Males in (a) Controls and (b) PFP. The interaction (Fig.3c) is obvious when contrasting (c) vs. (d), but is less obvious when contrasting (a) vs. (b), implying that qualitative comparisons of mean/SD trajectories can miss interactions.

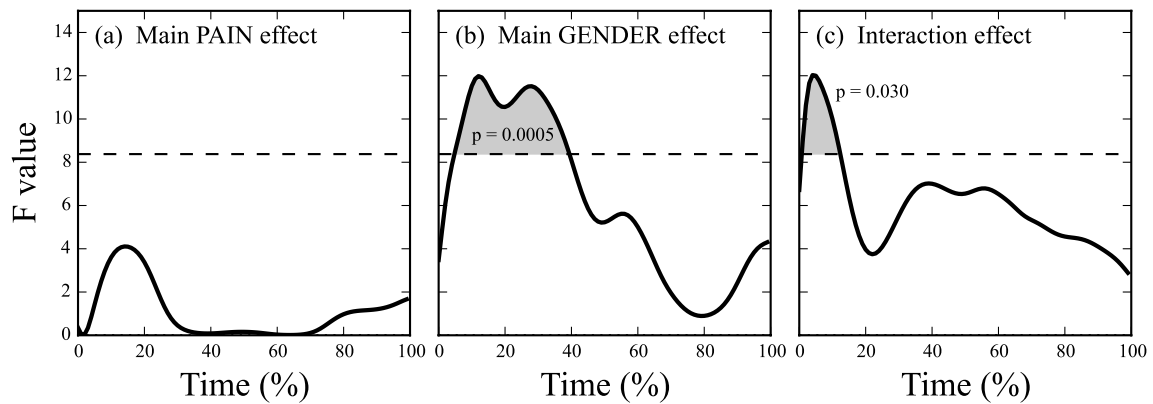


Figure 3. ANOVA results (full model, with interaction effects). Dotted horizontal lines depict the critical (parametric) RFT threshold at $\alpha=0.05$. Cluster-specific p values indicate the probability that Gaussian random trajectories would yield suprathreshold clusters of the same temporal extent.

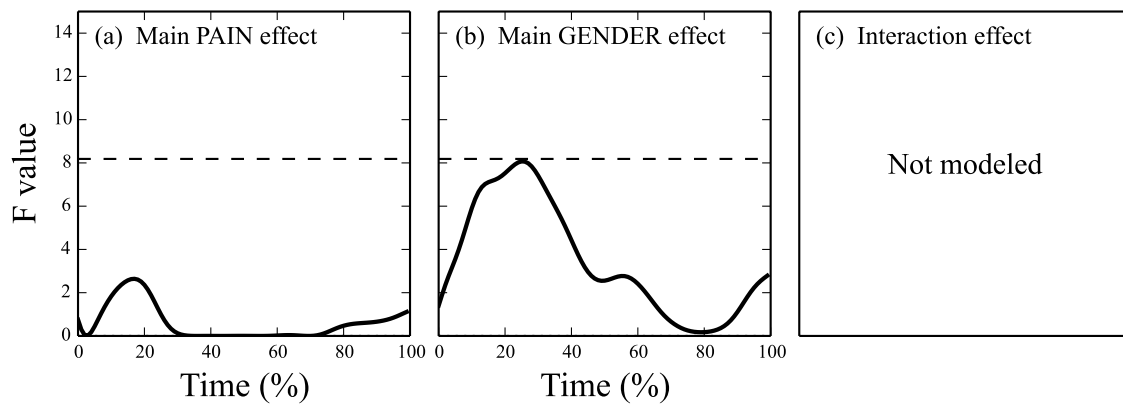


Figure 4. ANOVA results (main effects model, no interaction effect).

Appendix A. ANOVA computation overview

The experiment in the main manuscript consisted of two experimental factors: PAIN and GENDER, each with two levels: (control, PFP) and (female, male). As detailed in the main manuscript the response variable of interest was a (1×100) scalar trajectory, and there were a total of 41 responses: 8 control females, 7 control males, 16 PFP females and 10 PFP males.

We can model these data using a general linear model (GLM):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{A.1}$$

where \mathbf{Y} is a (41×100) matrix of the experimentally measured responses, \mathbf{X} is a (41×4) design matrix (Fig.A1), $\boldsymbol{\beta}$ is a (4×100) matrix of mean trajectories, and $\boldsymbol{\varepsilon}$ is a (41×100) matrix of residuals. Each column of \mathbf{X} corresponds to a PAIN-GENDER pair, and the j th row contains a single one and three zeros, with the one appearing in the column corresponding to the j th subject's pain condition and gender.

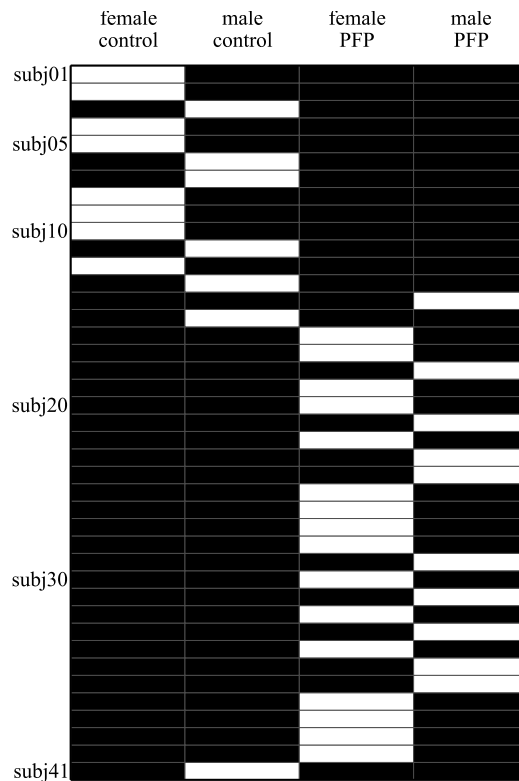


Figure A1: Experimental design matrix. White cells are ones and black cells are zeros.

The least-squares solution to Eqn.A.1 is:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (\text{A.2})$$

and the model's residuals are:

$$\hat{\epsilon} = \mathbf{Y} - \mathbf{X} \hat{\beta} \quad (\text{A.3})$$

The fitted $\hat{\beta}$ matrix is (4×100) , containing one mean trajectory for each column of \mathbf{X} . The residuals matrix $\hat{\epsilon}$ is (41×100) and contains the differences between the original data \mathbf{Y} and the relevant mean trajectory $\hat{\beta}$. From the perspective of Random Field Theory (RFT), ϵ are assumed to be smooth, Gaussian random fields.

The entire fitted model may be visualized as a pseudo-color plot (Fig.A2). Note that each row of \mathbf{Y} , $\hat{\beta}$ and $\hat{\epsilon}$ represents a single, temporally smooth trajectory.

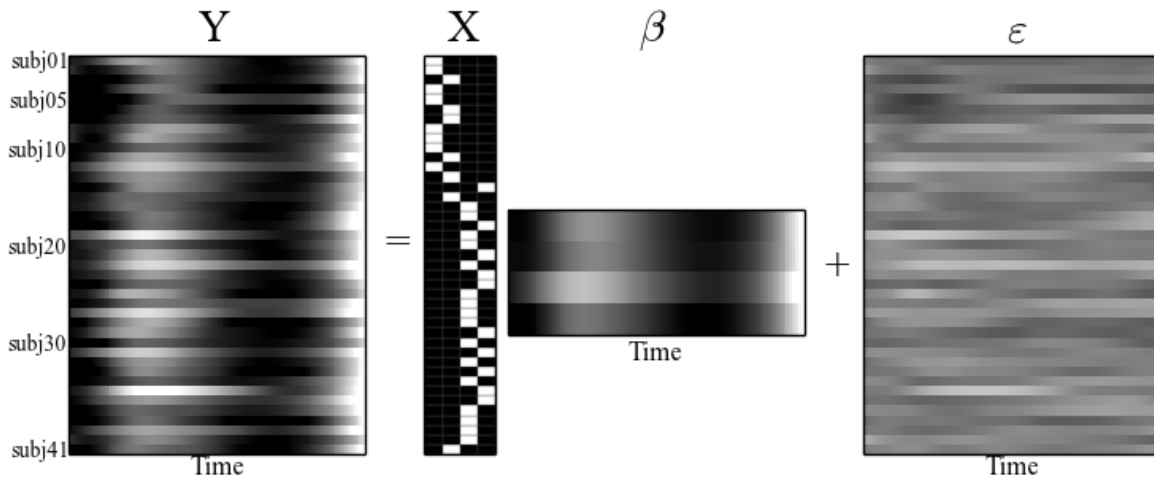


Figure A2: Statistical model (see Eqn.1). The time-normalized data (\mathbf{Y}) are modeled as a set of mean trajectories (β) about which each subject's trajectory varies smoothly (varepsilon). The design matrix (\mathbf{X}) is used to estimate the parameters (β) in a least-squares sense.

Since $\hat{\beta}$ and $\hat{\epsilon}$ respectively embody mean and variance trajectories, it is clear that they can be combined to form test statistics in general, and F statistics in particular. Unfortunately the computational details are somewhat complex, so we leave this discussion with a conceptual, generalized summary:

Arbitrary biomechanics experiments (e.g. t tests, regression, ANCOVA, etc.) can be modeled using \mathbf{X} , and when the data can be assembled into a single response matrix \mathbf{Y} , the model parameters and variances can be rapidly computed using Eqns.A.2&A.3. Then test statistic

fields can be constructed using combinations of $\hat{\beta}$ and $\hat{\epsilon}$, and we can conduct statistical inference by comparing our observed test statistic field to the behavior of Gaussian fields which are funnelled through the same experimental design \mathbf{X} .

Readers interested in additional computational details, and a more thorough treatment of ANOVA theory may wish to consult Christensen (1996) and Friston et al. (2007).

References

Christensen, R. 1996. *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer, New York.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and Penny, W. D. 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Elsevier/Academic Press, Amsterdam.

Appendix B. SPM vs. PCA

Statistical Parametric Mapping (SPM) and Principal Component Analysis (PCA) have emerged relatively recently in the Biomechanics literature. The primary difference between the two is that SPM is a hypothesis testing technique and PCA is a dimensionality reduction technique. This Appendix aims to explain this difference conceptually, as applicable to the analysis of experimental 1D trajectories.

To start, let us revisit the general linear model (Eqn.A.1), which is replicated here for convenience:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (\text{B.1})$$

The variables in this equation emerge, in chronological order, as follows:

1. \mathbf{X} : experimental design, set by an investigator prior to conducting an experiment.
2. \mathbf{Y} : experimental data, measured during the experiment.
3. $\boldsymbol{\beta}$: computed regression parameters, usually the least-squares map between the design \mathbf{X} and the data \mathbf{Y} .
4. $\boldsymbol{\varepsilon}$: computed model residuals, representing the experimental variability about the parameters $\boldsymbol{\beta}$.

Note that this model is applicable to all experimental designs including: t tests, regression, ANOVA (as detailed in Appendix A) and more complex designs like MANCOVA. For t tests and ANOVA the $\boldsymbol{\beta}$ parameters are mean trajectories (one per group), and we shall limit subsequent discussion to this case.

SPM and PCA are equivalent up until the end of Step #2: both involve analysis of \mathbf{Y} as measured during some experiment \mathbf{X} . SPM proceeds to Step #4, and then asks a conceptually simple question: *what is the probability that the effects embodied in $\boldsymbol{\beta}$ could be produced by random 1D trajectories like those embodied in $\boldsymbol{\varepsilon}$?* In a two-sample t test, for example, the two rows of $\boldsymbol{\beta}$ represent the two groups' mean trajectories, and those two trajectories are generally different. Difference itself is scientifically uninteresting because a variety of factors including measurement error ensure that mean trajectories are never precisely equivalent. Probabilities associated with trajectory differences are much more relevant: if random trajectories would frequently produce trajectory differences as large or larger than the observed mean trajectory differences, then the null hypothesis (of no difference) has successfully predicted the experimental result. On the other hand, if random trajectories would produce the observed difference relatively infrequently, then the null hypothesis failed to predict the experimental result and can be rejected. Formally, SPM quantifies such probabilities using Random Field Theory, which

analytically describes the frequency with which trajectory differences are expected to emerge when Gaussian random fields are routed through the experimental design \mathbf{X} . Like all 0D parametric hypothesis testing procedures, SPM regards the residual trajectories ε as independent and normally distributed, but these assumptions can easily be relaxed with non-parametric forms of SPM.

In contrast, PCA asks the following question: *what trajectories represent the most variance in \mathbf{Y} ?* Some of the resulting PCs may be similar to the sample means (β), but in general are different. Since PCA does not compute β directly, it effectively ignores the experimental design \mathbf{X} . This approach allows one to powerfully probe trends in \mathbf{Y} irrespective of \mathbf{X} , but by doing so one loses the ability to ask probabilistic questions which pertain to \mathbf{X} . The probabilistic meaning of PCA results only emerges when tested on independent datasets using one or more validation procedures, as described in the machine learning literature (Bishop C. M., 2007).

In summary, whereas SPM establishes a probabilistic link amongst all four model elements (Eqn.B.1), PCA instead analyzes the variability in \mathbf{Y} in isolation. The consequences are that SPM results generalize beyond the analyzed dataset, and that PCA results must be validated on independent datasets to establish generalizability. Most concisely: SPM is a hypothesis testing technique and PCA is a dimensionality reduction technique.

The practical implications are as follows: if one wishes to formally test *a priori* hypotheses regarding whole 1D trajectories, then SPM is a good choice. If, however, one wishes to describe the sources of variability within a particular dataset, then PCA is a good choice. The important scientific distinction is that, whereas SPM generates probability values corresponding to the given experimental dataset, PCA results can only adopt probabilistic meaning when validated on independent datasets. Interested readers may wish to consult machine learning textbooks (e.g. Bishop, 2007), which clarify the role of PCA and other dimensionality reduction techniques in the broader spectrum of probability computations.

References

Bishop C. M. 2007. *Pattern Recognition and Machine Learning*, Springer, New York.